

Stationary consistent equilibrium coalition structures constitute the recursive core*

László Á. Kóczy[†]

Abstract

We study coalitional games where the coalitional payoffs depend on the entire coalition structure. The recursive core (Kóczy, 2007) is a generalisation of the coalition structure core for such games.

We introduce a noncooperative, sequential coalition formation model and show that the set of equilibrium outcomes coincides with the recursive core. In order to extend past results to games that are not totally balanced (understood in this special setting) we introduce subgame-consistency that requires perfectness in relevant subgames only, while subgames that are never reached are ignored.

Subject classification: C71, C72

Keywords and phrases: partition function, externalities, implementation, recursive core, stationary perfect equilibrium, time consistent equilibrium

1 Introduction

Throughout its history the theory of coalitional games has mostly focussed on the study of games with *orthogonal coalitions*, that is, coalitions, which can be studied independently of each other. The most obvious example is the commonest form of a TU-game with a characteristic function that assigns a payoff

*The author acknowledges the support of the European Union (MEIF-CT-2004-011537, PERG-GA-2008-230879) and OTKA - Hungarian Fund for Scientific Research (NF-72610).

[†]Budapest Tech, Faculty of Economics, Tavaszmező 15-17., H-1084 Budapest, Hungary and Maastricht University, Department of Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands, koczy.laszlo@kgk.bmf.hu

to a coalition disregarding other players and coalitions. When we look at the usual interpretations of coalitions, be those trading blocks (Yi, 1996), trusts (Bloch, 1995) or international environmental agreements (Funaki and Yamato, 1999; Eyckmans and Tulkens, 2003), the orthogonality assumption is difficult to maintain; we believe it is the exception rather than the rule that coalitions can be studied independently of each other.

Since the seminal paper of Thrall and Lucas (1963) introducing the partition function form numerous cooperative approaches and solution concepts have been proposed to solve games with externalities, but in the absence of an implementation by non-cooperative equilibria these remain interesting heuristics (Chander and Tulkens, 1995; Ray and Vohra, 1997; Hyndman and Ray, 2007). For games with orthogonal coalitions the implementation of cooperative solution concepts, such as the core has an extensive literature (Chatterjee et al., 1993; Lagunoff, 1994; Perry and Reny, 1994), but these results do not directly generalise to games with externalities. In this domain Huang and Sjöström (2006) and Kóczy (2009) have provided partial results that are limited to games with non-empty cores in all subgames, or, in terms of sequential coalition formation games: to games with stationary perfect equilibria. It turns out that perfectness is a very demanding condition and the implementation might fail even for simple TU games. We therefore introduce a generalisation, subgame-consistency, and show that the set of partitions formed under the resulting equilibria coincides with the recursive core.

Subgame-consistency is a weaker concept than subgame-perfectness, but more demanding than time-consistency (Kydland and Prescott, 1977). If we define each of these concepts in corresponding sets of subgames, for subgame-perfectness all subgames are relevant, while for time-consistency only the subgames on the equilibrium path. In particular subgame-perfect equilibria are also subgame-consistent and subgame-consistent equilibria are also time-consistent. Moreover stationary perfect equilibria are stationary consistent. For more on the relation of subgame-perfect and time-consistent strategies see Fershtman (1989) and Asilis (1995).

The structure of the paper is as follows. After this introduction a long second section follows introducing both the cooperative and noncooperative theories to

solve games in partition function form, we introduce the notation and simple terminology we are going to use. We present the cooperative solution, namely the recursive core and similarly the noncooperative coalition formation game and its equilibria. A novel equilibrium concept, *subgame consistency* and the corresponding notion of *relevant subgame* are also introduced here. We state and prove our main result in the third section. The paper ends with a brief conclusion.

2 Preliminaries

Let N denote the set of players. Subsets are called *coalitions*. A partition \mathcal{S} of S is a splitting of S into disjoint coalitions. $\Pi(S)$ denotes the set of partitions of S . In general we use capital and calligraphic letters to denote a set and its partition (the set of players N being an exception), indexed capital letters are elements of the partition. We write $i \in \mathcal{S}$ if there exists S_k such that $i \in S_k \in \mathcal{S}$ and if $i \in \mathcal{S}$ we write $\mathcal{S}(i)$ for the coalition embedded in \mathcal{S} containing i .

The game (N, V) is given by the player set N and a *partition function* (Thrall and Lucas, 1963) $V : \Pi(N) \rightarrow (2^N \rightarrow \mathbb{R})$, where $V(S_i, \mathcal{S})$ denotes the payoff for coalition S_i in case partition \mathcal{S} forms. For vectors $x, y \in \mathbb{R}^N$ we write x_S for the restriction to the set S and $x_S > y_S$ if $x_i \geq y_i$ for all $i \in S \subset N$ and there exists $j \in S$ such that $x_j > y_j$.

The pair $\omega = (x, \mathcal{P})$ consisting of a payoff vector $x \in \mathbb{R}^N$ and a partition $\mathcal{P} \in \Pi(N)$ is a *payoff configuration* (or *outcome*) if $\sum_{i \in S} x_i = V(S, \mathcal{P})$ for all $S \in \mathcal{P}$. The set of outcomes of game (N, V) is denoted $\Omega(N, V)$.

Our main result is the equivalence of the partitions produced by certain noncooperative coalition formation game and a cooperative solution concept. In the following we spell out these approaches.

2.1 Recursive core

The first model is a cooperative solution concept, a generalisation of the core to games in partition function form. The core is defined in terms of deviations, but unlike in games with orthogonal coalitions, in games with externalities the profitability of a deviation can only be determined once the partition of the

remaining, *residual* players is also known, or at least some assumption is formulated about their behaviour. While most of the approaches (see Kóczy, 2007, for further references) tried to get rid of the externalities and solve the game as a characteristic function form game, Huang and Sjöström (2003) and (Kóczy, 2007) assume that these residual players play a residual game that is a game on its own and thus can be solved using the same concept. Once the solution of this game is known, we know which partition is formed, and then it is also possible to tell the deviating players' payoffs. If this partition is not unique (or not determined, in case the residual core is empty) Kóczy (2007) considers optimistic and pessimistic scenarios depending on the deviating players' expectations regarding these alternatives. Our results will apply to the pessimistic case, so only this version of the definition is given.

First we introduce *residual games* and then the recursive core:

Definition 1 (Residual Game). Let (N, V) be a game and consider the set $L \subsetneq N$ of live players. Assume $K = N \setminus L$ have committed to form partition \mathcal{K} . Then the *residual game* $(L, V^{\mathcal{K}})$ is the partition function form game over the player set L and with the partition function $V^{\mathcal{K}} : \Pi_L \rightarrow (2^L \rightarrow \mathbb{R})$, where

$$V^{\mathcal{K}}(C, \mathcal{L}) = V(C, \mathcal{L} \cup \mathcal{K}) \quad \forall C, \mathcal{L} : C \in \mathcal{L} \in \Pi_L. \quad (2.1)$$

The residual game is *derived* from the original game using the partition \mathcal{K} , but it is a partition function game on its own. So *if* we use the core to solve (N, V) , we must also use it to solve $(L, V^{\mathcal{K}})$: Deviating coalitions must expect a residual core outcome to form. Should the core be empty this solution does not present a selection of the outcomes, and all possible responses must be considered. Even if the residual core is non-empty it may contain payoff configurations with different partitions. This gives rise to the following definition.

Definition 2 (Recursive core). Let (N, V) be a partition function form game.

1. *Trivial game*. The *core* of $(\{1\}, V)$ is the only outcome with the trivial partition: $C(\{1\}, V) = \{(V(1, (1)), (1))\}$.
2. *Inductive assumption*: The core $C(N, V)$ has been defined for all games

with $|N| < k$ players. The *assumption about game* (N, V) is

$$A(N, V) = \begin{cases} C(N, V) & \text{if } C(N, V) \neq \emptyset \\ \Omega(N, V) & \text{otherwise.} \end{cases}$$

3. *Dominance.* The outcome (x, \mathcal{P}) is *dominated via the coalition K forming partition \mathcal{K}* if for all assumptions $(y_L, \mathcal{L}) \in A(L, V^{\mathcal{K}})$ of the remaining set of players $L = N \setminus K$ there exists an outcome $((y_K, y_L), \mathcal{K} \cup \mathcal{L}) \in \Omega(N, V)$ such that $y_K > x_K$.

The outcome (x, \mathcal{P}) is *dominated* if it is dominated via a coalition.

4. *Core.* The *core*, denoted $C(N, V)$, is the set of undominated outcomes.

A partition is only dominated via a coalition if the deviation of this coalition (as a partition) is profitable for *every* residual (core) partition. When the residual core is empty, we have no information about the solution of the residual game, so we assume that any reaction is possible. As such, we do not, for instance, exclude inefficient partitions – just as the sequential game will be free from such limitations in Equation 2.3. Our results, however, generalise to such modifications – as long as they are introduced in both models. For a general discussion of the properties of the recursive core see Kóczy (2007).

In the following we simply refer to the recursive core as *core* and to the (recursive) core of a residual game as *residual core*.

2.2 Sequential coalition formation

In the following we describe the noncooperative coalition formation game that, together with an appropriate equilibrium concept will implement the recursive core. Our setup is closest to the models of (Bloch, 1996) and Perry and Reny (1994).

We begin with a brief, informal description. The purpose of the game is to form coalitions as this is the only way players can collect payoffs. Coalitions can form with unanimous agreement of the members: a member proposes a coalition and the rest one-by-one accept. Such agreements must also specify the distribution of the coalitional payoffs, which, in turn, are only determined once the game ends and the complete partition forms. Following Huang and Sjöström

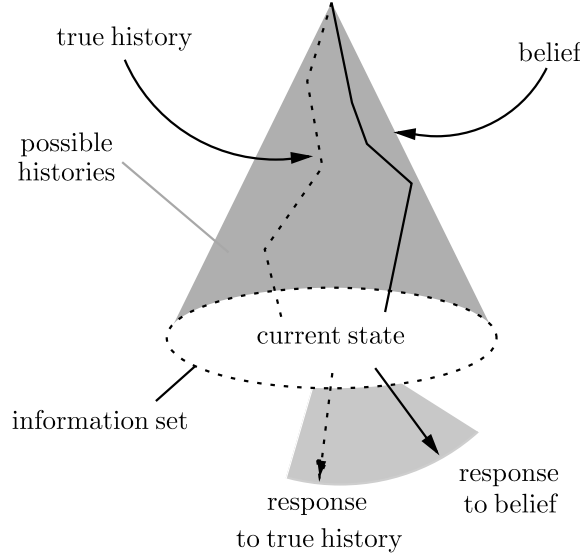


Figure 1: Stationary decision with beliefs

(2006) we assume that a proposal also specifies the percentual distribution of the payoff among the members of the coalition. For efficiency reasons we allow for the simultaneous formation of coalitions (See also Kóczy, 2009). When a coalition forms, its members leave the game.

We are interested in self-confirming strategies of this game. It is common to study stationary-perfect equilibria, but here perfectness turns out to be too strong a condition and therefore we introduce the somewhat weaker notion of subgame-consistency that requires the perfectness property only in *relevant* subgames. We insist on stationarity, but allow players to form *beliefs* about history: their actions may thus depend not only on the status quo, but also on the common belief, decided by nature.

The decision at each node can therefore be described as on Figure 1. At a node there may be several possible moves each of which are optimal for a different history. Also here, the player should follow the dotted arrow, but, due to stationarity the decision can only be conditional on the belief. The possibility, however is there that the right action is chosen, as the belief can possibly coincide with the true history.

In the following we formalise the model.

The sequential coalition formation game (N, V) is defined over the same

player set N and the same partition function V , but the game is played in an entirely different way. Without loss of generality we assume $0 < \min_{\mathcal{P}, S} V(S, \mathcal{P})$ therefore staying in the game forever is never optimal.

1. Initially all players are active and no proposals have been made.
2. Player i makes a *proposal* to an active subset $S \ni i$ of the players specifying a partition \mathcal{S} as well as a distribution of the coalitional values.
3. If it is attractive, players in S accept the proposal one-by-one.
4. When all players have accepted, the proposed coalitions form and leave the game.
5. The coalitions that have left receive some payment based on what they have already earned, that is, the minimal payment for these coalitions in any embedding partition taking the exited coalitions given.
6. The *belief* about the order of exits of departed coalitions is updated by nature; the game continues with the remaining players with Step 2.
7. If a proposal is not attractive, the invited players do not accept it and another proposal is made. Return to Step 2.¹
8. If all players have left, the game ends.

Proposals A proposal $p = (\mathcal{T}, w) \in \Pi_T \times [0, 1]^T$ by player i is offered to a set of players $T \ni i$, where $T \subseteq N \setminus K$ specifying a partition $\mathcal{T} \in \Pi_T$ and a distribution w of coalitional payoffs in each of the coalitions in \mathcal{T} such that $\sum_{i \in T_i} w_i = 1$ for all $T_i \in \mathcal{T}$. The vector w specifies the *share* of the payoff a particular player will receive, the actual payment is only known once the payoff of the coalition can be determined. Specifying individual payoffs adds only complexity to the model (Huang and Sjöström, 2006).

The set of proposals available to i are denoted P_i while P collects all possible proposals.

¹Here we use the assumption that there are plenty of opportunities to accept a proposal so if a player does not, but allows another one to make another proposal, then he is essentially rejecting it. Bloch (1996) only allows the rejecting player to make a proposal, but with this assumption our setup is essentially the same.

History The game is specified in an extensive form, where decisions are made at each node. History, denoted h^t at time t encompasses the entire activity log of the game including proposals made, acceptances if any, which coalitions have left etc.

Time's passing has no relevance here, so we drop the reference in the notation. We can, however say that if $h_1 \subset h_2$, then h_1 happened before h_2 .

History has more data we will ever need, among others it contains.

- the set of players $K(h) = \bigcup_{S \in \mathcal{K}(h)} S \subset N$ who have already left the game forming partition $\mathcal{K}(h) \in \Pi_{K(h)}$,
- the set of feasible proposals $P(h) = \{(\mathcal{T}, w) \in P \mid \mathcal{T} \subseteq N \setminus K(h)\}$ and $P_i(h) = P(h) \cap P_i$,
- the current proposal $p(h) = (\mathcal{T}(h), w(h)) \in P(h)$,
- the distribution rule for the quit players $w(h) \in \mathbb{R}^N$, where we set $w_i(h) = \frac{1}{|\mathcal{I}|}$ for all $i \notin K$,
- the set of players $A(h) \subset \mathcal{T}(h)$ who have already accepted the proposal,
- and finally $\beta(h)$ the *belief* at node h .

The set of histories is denoted by H .

When a history h has been reached, all future histories can only be extensions of h . The set of such feasible histories is denoted $H^h = \{h' \in H \mid h \subseteq h'\}$.

Strategies *Strategy* σ_i of player i is a mapping from H to his actions set:

$$\sigma_i(h) \in P_i(h) \cup \{\text{accept}, \text{wait}\}. \quad (2.2)$$

We assume that making or accepting a proposal is preceded and followed by a nonempty open passive interval of time. This is to ensure that other players have a chance to react. Further, we assume

1. Initially there is no active proposal so “accept” is the same as “wait”.
2. The same applies if i accepts a proposal (\mathcal{T}, w) , while $i \notin \mathcal{T}$.
3. Actions of departed players are ignored.

4. A new proposal cancels the previous proposal: if it was not accepted, by our assumption this is due to lack of interest not shortage of time. Here the question whether a race-to-react could realise in some situations. Fortunately the answer is no, in equilibrium this will not happen, but to see this we first must specify payoff (the incentives to play the game) and the equilibrium concept.

We denote the restriction of σ to a subgame corresponding to history h by σ^h .

Beliefs In games with externalities stationary information is not sufficient to achieve stable outcomes. Indeed a departed player's payoff is partly determined by the actions of the remaining players allowing the latter to punish deviations. A strategy will typically specify punishment strategies off equilibrium, but whom should one punish depends on the history, that is not available when stationary strategies are selected. Instead players form a common belief on whom should be punished. The belief is common to all players, and is decided by nature and is updated each time a player leaves a game.

For all h , the belief $\beta(h)$ at h is a subset of the departed coalitions, $\beta(h) \subseteq \mathcal{K}(h)$. As for strategies, the restriction of β to subgame h is denoted by β^h . The set of beliefs is denoted B , the set of restrictions to h by B^h .

In equilibrium the outcome of the game will not depend on beliefs, but to make the influence of recollections more explicit in general, the outcome resulting from the strategy profile σ and the recollection-function β can be written as $\omega(\sigma, \beta)$. Kreps and Wilson (1982) use expectations to aggregate results from different beliefs, we use the conservatism of the players: They focus on the worst outcomes, essentially trying to minimise loss.

Payoffs Not all players will necessarily leave the game – if so, the departed players form a coalition structure that is not a partition and therefore the coalitional payoffs are ill-defined. Assuming that players are careful, conservative and thus always look out for the worst case and with a slight abuse of notation we generalise the payoff function for such “incomplete partitions”.

$$V(S, \mathcal{P}(\sigma, \beta)) = \begin{cases} \min_{\mathcal{P} \supset \mathcal{P}(\sigma, \beta)} V(S, \mathcal{P}) & S \in \mathcal{P}(\sigma, \beta) \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

In addition to the coalitional payoffs, the strategies also determine the *individual* ones. Let $x_i(\sigma, \beta)$ denote the payoff of player i in case the strategy profile σ is played and β is the belief function. Formally

$$x_i(\sigma, \beta) = w_i(\sigma, \beta) V(\mathcal{P}(\sigma, \beta, i), \mathcal{P}(\sigma, \beta)), \text{ or} \quad (2.4)$$

$$x_i(\sigma) = \min_{\beta' \in B^h} w_i(\sigma, \beta') V(\mathcal{P}(\sigma, \beta', i), \mathcal{P}(\sigma, \beta')), \quad (2.5)$$

as a function of strategies only.²

Before we proceed to study equilibria we introduce some additional notation. At history h the continuation payoff for i using strategy profile σ is

$$x_i(\sigma, h) = \begin{cases} \min_{\beta^h \in B^h} w_i(h) V(\mathcal{K}(h, i), \mathcal{K}(h) \cup \mathcal{P}(\sigma^h, \beta^h)) & \text{if } i \in K(h) \\ \min_{\beta^h \in B^h} w_i(\sigma^h, \beta^h) V(\mathcal{P}(\sigma^h, \beta^h, i), \mathcal{K}(h) \cup \mathcal{P}(\sigma^h, \beta^h)) & \text{otherwise} \end{cases} \quad (2.6)$$

Now suppose that at history h there is a proposal $p = (\mathcal{D}, w)$. For a player $i \in \mathcal{D}$, should the proposal be accepted the payoff becomes

$$x_i(\sigma, \mathcal{D}, w, h) = \min_{\beta^h \in B^h} w_i V(\mathcal{D}(i), \mathcal{K}(h) \cup \mathcal{D} \cup \mathcal{P}(\{\text{accept}\} \cup \sigma_{-i}^h, \beta^h)). \quad (2.7)$$

Finally we introduce the following notation: Suppose $x(\beta)$ and $y(\beta)$ are payoff vectors. We say that x is *larger* than y for all $\beta \in B$ and write $x(B) > y(B)$ if $x(\beta) \geq y(\beta)$ for all $\beta \in B$ and there exists $\beta \in B$ such that $x(\beta) > y(\beta)$.

Equilibria Now that we have specified the available strategies (actions), the resulting payoffs (incentives) we can focus on the outcomes of the coalition formation game. We hope to answer two questions simultaneously: (i) which

²Bloch (1996) considers optimistic players: $v_i(\mathcal{P}(\sigma)) = \max_{\mathcal{P} \supset \mathcal{K}} v_i(\mathcal{P})$ for $i \in \mathcal{P}(\sigma)$. However, a deviation (a concept we formalise later) is profitable if it is weakly profitable to all players. Suppose this deviation creates a subgame where any of the partitions might form. Optimistic players expect the best: a partition beneficial to the deviation will form. Bloch's players' optimism goes further, they individually hope the best: a deviation may appear profitable even if for *every single* possible reaction someone is worse off. Pessimism is consistent in this sense: a deviation \mathcal{K} is profitable for K if and only if it is profitable for each player in K and for each possible partition.

coalitions will form (ii) how are coalitional payoffs distributed. We look for strategies that do not need revisions, but are final already as the game starts and for strategies that are stationary, that is, do not depend on time, but only on the current state of the game. In nonstationary strategies the set of equilibria may be too inclusive; for a discussion of folk-theorem-like results see Muthoo (1990, 1995); Perry and Reny (1994); Osborne and Rubinstein (1990).

Definition 3. A strategy is *stationary* if it does not depend on history, if for all i and for all histories $h, h' \in H$ such that $\mathcal{K}(h) = \mathcal{K}(h')$ and $p(h) = p(h')$ we have $\sigma_i(h) = \sigma_i(h')$.

Such strategies only depend on the current state $s = (\mathcal{K}, p)$, a pair consisting of the partition \mathcal{K} of departed players and the ongoing proposal $p \in P^i(h)$.

Now recall that players are conservative and only go for certain profits: If different beliefs lead to different subsequent actions from the other players, a deviation may or may not be profitable under all such scenarios.

Definition 4. The strategy profile σ^* is a subgame-perfect equilibrium (with beliefs) if for all $i \in N$, for all $h \in H_i$ and for all strategies σ_i we have

$$x_i(\sigma_i^{*h}, \sigma_{-i}^{*h}, B^h) > x_i(\sigma_i^h, \sigma_{-i}^{*h}, B^h). \quad (2.8)$$

Definition 5. A *stationary perfect equilibrium* σ^* is a strategy profile that is both subgame-perfect and stationary.

The set of stationary perfect equilibrium partitions coincide with the recursive core (Kóczy, 2009) (for games with nonempty residual cores). This equivalence result predicts that games with empty residual cores do not have stationary perfect equilibria.

Bloch (1996) presents a 3-player example, where player 1 would like to form a coalition with 2, 2 with 3, 3 with 1. This game does not have stationary-perfect equilibria. Since residual games are also partition function form games, the smallest residual game for which the corresponding subgame of the sequential game has no stationary strategies has an empty core. By a sufficiently large payoff for the grand coalition the core of the original game is nevertheless empty. Perfectness only holds globally, that is, if the tiniest subgame fails to have stationary perfect equilibria this imperfection spreads to the entire game. On

the other hand, just as the recursive core may be non-empty even if the game has empty residual cores, with a weaker concept of perfection we may retain an essentially perfect behaviour in the corresponding sequential coalition formation games, too.

Time-consistency (Kydland and Prescott, 1977) merely requires that the equilibrium strategy does not need revision and thus will naturally be unaffected by empty cores in subgames elsewhere. Clearly, most subgames are never reached so it is superfluous to insist on this property everywhere, on the other hand time-consistency does not check deviations carefully enough. *Subgame-consistency*, that we introduce soon is the right compromise and in the following we study strategies where the perfectness/consistency criterion is not checked per se for every subgame, but is only required in *relevant* subgames.

Definition 6. For a strategy profile σ and belief β a subgame at history h is *relevant* if

1. h is the original game ($\mathcal{K}(h) = p(h) = \emptyset$), or
2. there exists a modification σ' and a belief β^h such that
 - (a) σ and σ' differ in a single action in history h , resulting in the set D forming partition $\mathcal{D} = \mathcal{K}(h) \setminus \mathcal{P}(\sigma, \beta)$ leaving the game,
 - (b) $\mathcal{K}(h) \subseteq \mathcal{P}(\sigma', \beta^h)$, and
 - (c) $x_D(\sigma', \beta^h) > x_D(\sigma, \beta^h)$, or
3. it is a relevant subgame of a relevant subgame.

The first case is trivial. In Case 2 we consider an elementary irreversible deviation (the wrong set D of players or the wrong partition \mathcal{D} exits), that is, nevertheless profitable for D under some belief. Finally if some elementary deviation results in a subgame this subgame is relevant and of course after the departure of some of the original players, it is a game on its own, and is therefore evaluated the very same way: we want to check relevant subgames.

Definition 7. The strategy profile σ^* is a *subgame-consistent equilibrium*³ if

³In this equilibrium concept perfectness is only required in the neighbourhood of the equilibrium strategies and so the name quasi-perfect equilibrium would be more appropriate.

- for all players $i \in N$, for all histories h and for all strategies σ_i

$$x_i(\sigma_i^{*h}, \sigma_{-i}^{*h}, B^h) > x_i(\sigma_i^h, \sigma_{-i}^{*h}, B^h) \quad (2.9)$$

- restrictions to subgames relevant for σ are also subgame-consistent.

Definition 8. A *stationary consistent equilibrium* σ^* is a strategy profile that is both subgame-consistent and stationary.

We denote the set of stationary consistent equilibria by $\text{SCE}(N, V)$ and outcomes resulting from playing such equilibrium strategies by $\Omega^*(N, V)$.

3 Results

Theorem 1. Let (N, V) be a partition function form game. Then its recursive core $C(N, V)$ coincides with the set $\Omega^*(N, V)$ of outcomes supported by stationary consistent equilibrium strategy profiles.

The rest of this section is devoted to the inductive proof of this theorem. As the proof is long, we break it into a number of propositions and finally present a summary of these results. The first proposition requires no proof:

Proposition 2. Let $(\{1\}, V)$ be a trivial, single-player partition function form game. Then $C(\{1\}, V) = \Omega^*(\{1\}, V)$.

Now *assume* that Theorem 1 holds for all games with less than k players. In the following we extend it to games with k players. In order to show $\Omega^*(N, V) = C(N, V)$, first we show $\Omega^*(N, V) \subseteq C(N, V)$ then $\Omega^*(N, V) \supseteq C(N, V)$.

Lemma 3. If Theorem 1 holds for all games with up to $k-1$ players, $\Omega^*(N, V) \subseteq C(N, V)$ for all k -player games.

Proof. If $\Omega^*(N, V) = \emptyset$ the result is trivial, otherwise there exists a SCE σ producing $\omega(\sigma, \beta) = (x(\sigma, \beta), \mathcal{P}(\sigma, \beta)) \in \Omega^*(N, V)$ for some belief-function β . In particular, we assume that $\omega(\sigma, \beta) \notin C(N, V)$ and prove contradiction.

Unfortunately that term is already taken and quasi-perfect equilibria (van Damme, 1984) are, however not related to our concept, in fact even their relation to subgame perfect equilibria is not well defined. Our concept is a weakening of subgame-perfectness.

By this assumption there exists a profitable deviation \mathcal{D} by some set D of players. The induced subgame has fewer players so the inductive assumption can be applied. In the sequential game the deviation at h is expressed by the strategy profile σ' against the original strategy profile σ , where $\sigma'(h') = \sigma(h')$ for all $h' \subset h$. We discuss three cases.

Case 1. The induced subgame with $\mathcal{K}(h) = \mathcal{D}$, $p(h) = \emptyset$ is not relevant. Then for all $\sigma_{-i}(h)$ there exists $i \in D$ and β such that $x_i(\sigma', \beta) < x_i(\sigma, \beta)$ – thus the deviation cannot be profitable in the cooperative game; contradiction.

Case 2. The resulting subgame is relevant, the core of the corresponding residual subgame is empty. Then $V(D, \mathcal{D} \cup \mathcal{P}_{N \setminus D}) > \sum_{i \in S} x_i(\sigma, \beta)$ for all $\mathcal{P}_{N \setminus D}$. Since $V(D, \mathcal{D} \cup \mathcal{P}_{N \setminus D}) = \min_{\beta^h \in B^h} \sum_{i \in S} x_i(\sigma', \beta)$ a player in D should immediately propose \mathcal{D} . By subgame consistency all in D will accept. Therefore σ is not a stationary consistent equilibrium, moreover the outcome $\omega(\sigma, \beta)$ cannot be supported by other equilibria either. Contradiction.

Case 3. The induced subgame is relevant and the core of the corresponding residual subgame is not empty. Since σ is a SCE its restriction σ^h to this relevant subgame (where $\mathcal{K}(h) \supset \mathcal{D}$ and $p(h) = \emptyset$) is stationary consistent, too. Moreover the deviation from σ to form \mathcal{D} is not profitable, therefore

$$x_D(\sigma^h, B^h) > x_D(\sigma'^h, B^h) \quad (3.1)$$

On the other hand, by the inductive assumption,

$$\omega(\sigma^h, \beta^h) \in C(L(h), V^{\mathcal{D}}) \quad \forall \beta(h). \quad (3.2)$$

This, however, implies that the deviation \mathcal{D} is not profitable in the cooperative game; contradiction.

We have discussed all cases, and found the assumptions contradicting. Therefore $\omega(\sigma, \beta) \in C(N, V)$. \square

Punishment strategy Before we move on to our next lemma, we introduce what we call the *punishment strategy*. A strategy-profile can only form an equilibrium if it specifies a “response” to each deviation that deems these deviations unprofitable. In the recursive core a deviation is only profitable if it represents an improvement in the payoffs for all residual assumptions. In the sequential

game, however, primary deviations can be punished, but due to stationarity, but with multiple departed coalitions finding the right punishment is difficult.

Consider the following example. In equilibrium players obtain a payoff x . Suppose coalitions A and B have left the game and they do not form a subset of the equilibrium partition (for simplicity: none of them do), therefore someone has deviated. If A deviated first $N \setminus A$ (including B) should stop this deviation producing payoff y^A such that $y_A^A < x_A$. When B deviates, too, the remaining $N \setminus A \setminus B$ should also choose an action to get a payoff z^B such that $z_B^B < y_B^A$. Consequently B does not deviate, y^A forms which is bad for A , hence A does not deviate and the equilibrium is preserved.

If, however $N \setminus A \setminus B$ are misinformed and think B deviated first. They want response y^B , which A did not comply with thus A must be punished by z^A , where $y_B^B < x_B$ and $z_A^A < y_A^B$. If $z_B^A > y_B^B$ and $z_A^A > x_A$ the response does not work. Fortunately, when the deviations are made it is not yet known what will be the belief of $N \setminus A \setminus B$: it can also be the correct history.

In the following we specify the punishment strategy to a deviation knowing that some other coalitions left, too. We assume that \mathcal{K} has already left the game, but \mathcal{D} was (or at least we think it was) the last to exit. Consider a proposal $p = (\mathcal{D}, w)$. In the partition function form game $(N \setminus K \cup \mathcal{D}, V^{\mathcal{K} \setminus \mathcal{D}})$ the partition \mathcal{D} , as a deviation, defines a residual game $(N \setminus K, V^{\mathcal{K}})$. We discuss two cases based on the emptiness of the core of this residual game.

If the residual core is *not empty* a “punishment strategy” to \mathcal{D} is $(\tilde{x}^{\mathcal{D}|\mathcal{K}}, \tilde{\mathcal{P}}^{\mathcal{D}}) \in C(N \setminus K, V^{\mathcal{K}})$ ensuring that the deviation \mathcal{D} is not profitable. That is, $\tilde{\mathcal{P}}^{\mathcal{D}|\mathcal{K}}$ satisfies⁴

$$\exists S \in \mathcal{D} : \quad V(S, \mathcal{K} \cup \tilde{\mathcal{P}}^{\mathcal{D}|\mathcal{K}}) < \sum_{i \in S} \tilde{x}_i, \quad \text{or} \quad (3.3)$$

$$\forall S \in \mathcal{D} : \quad V(S, \mathcal{K} \cup \tilde{\mathcal{P}}^{\mathcal{D}|\mathcal{K}}) = \sum_{i \in S} \tilde{x}_i. \quad (3.4)$$

Since $(\tilde{x}, \tilde{\mathcal{P}}) \in C(N, V)$ such a $\tilde{\mathcal{P}}^{\mathcal{D}|\mathcal{K}}$ exists for all deviations \mathcal{D} . Without loss of generality let $\tilde{w}_i^{\mathcal{D}|\mathcal{K}} = \frac{\tilde{x}_i^{\mathcal{D}|\mathcal{K}}}{\sum_{j \in \tilde{\mathcal{P}}^{\mathcal{D}|\mathcal{K}}(i)} \tilde{x}_j^{\mathcal{D}|\mathcal{K}}}$.

If the residual core is *empty* we observe that in order for a deviation to be profitable it must be profitable for *all* residual partitions. Since $(\tilde{x}, \tilde{\mathcal{P}}) \in$

⁴Observe that from the point of view of externalities only the residual partitions matter, and therefore we ignore payoffs.

$C(N, V)$, in the partition function form game the deviation is not profitable guaranteeing the existence of a residual partition $\tilde{\mathcal{P}}^{\mathcal{D}|\mathcal{K}} \in \Pi_{N \setminus D}$ satisfying Condition 3.3 or Condition 3.4. Here $\tilde{w}^{\mathcal{D}|\mathcal{K}}$ can be chosen arbitrarily, so let $\tilde{w}_i^{\mathcal{D}|\mathcal{K}} = \frac{1}{|\tilde{\mathcal{P}}^{\mathcal{D}|\mathcal{K}}(i)|}$.

Lemma 4. *If Theorem 1 holds for all games with less than k players, then $\Omega^*(N, V) \supseteq C(N, V)$ for all k -player games (N, V) .*

Proof. The proof is inspired by that of Bloch (1996, Proposition 3.2) in part, and is by construction. We show that if $(\tilde{x}, \tilde{\mathcal{P}}) \in C(N, V)$ there exists a stationary consistent strategy profile $\tilde{\sigma}$ such that for all β we have $\mathcal{P}(\tilde{\sigma}, \beta) = \tilde{\mathcal{P}}$ and $x(\tilde{\sigma}, \beta) = \tilde{x}$. Let $\tilde{w} = \frac{\tilde{x}_i}{\sum_{j \in \tilde{\mathcal{P}}(i)} \tilde{x}_j}$.

In the following we define the stationary strategy $\tilde{\sigma}_i$ for player i . Due to stationarity it is sufficient to specify the strategy for each triple $(\mathcal{K}, p, \mathcal{B})$ consisting of the partition of players who have already quit, the current proposal and the *current* belief (thus $\mathcal{B} \subseteq \mathcal{K}$).

$$\tilde{\sigma}_i(\mathcal{K}, \mathcal{T}, w, \mathcal{B}) = \begin{cases} (\tilde{w}, \tilde{\mathcal{P}}) & \text{if } \mathcal{K} = \mathcal{T} = \emptyset \\ (\tilde{w}^{\mathcal{B}|\mathcal{K}}, \tilde{\mathcal{P}}^{\mathcal{B}|\mathcal{K}}) & \text{if } \mathcal{T} = \emptyset, \text{ but } \mathcal{K} \neq \emptyset \\ \text{accept} & \text{if } x_i(\tilde{\sigma}, \mathcal{T}, w, (\mathcal{K}, \mathcal{B})) > x_i(\tilde{\sigma}, (\mathcal{K}, \mathcal{B})) \\ \text{wait} & \text{otherwise.} \end{cases} \quad (3.5)$$

In equilibrium $\mathcal{P}(\tilde{\sigma}) = \tilde{\mathcal{P}}$ and the strategy is stationary by construction so we only need to verify subgame-consistency. We show this by induction. As subgame-consistency holds for a trivial game we may assume that it holds for all games of size less than $|N|$.

Now consider game (N, V) and observe that if K departed to form \mathcal{K} the subgame is simply a coalition formation game with less players. We discuss two cases based on the emptiness of the residual core.

1. If the residual core is not empty, the proposed strategy exhibits the same similarity property: in equilibrium the core partition is proposed and accepted, while residual cores form off-equilibrium.

The inductive assumption then ensures that the off-equilibrium path is subgame-consistent so we only need to check whether a deviation \mathcal{T} is ever accepted.

This deviation corresponds to a deviation in the partition function game. Since $(\tilde{x}, \tilde{\mathcal{P}}) \in C(N, V)$, by the construction of $(\tilde{w}^{\mathcal{B}|\mathcal{K}}, \tilde{\mathcal{P}}^{\mathcal{B}|\mathcal{K}})$ we know that for some \mathcal{B} there exists a player in T for whom the deviation \mathcal{T} is not profitable. Given the pessimism of the players, this is sufficient to deter this player from accepting the proposal to deviate.

2. If the residual core is empty, the deviation is not profitable irrespective of the residual partition that forms, the subgame is not relevant, and therefore the second condition for subgame-consistency is satisfied.

The emptiness of the residual core, by our assumption, also implies that there are no stationary consistent equilibrium strategy profiles. In the absence of such strategy profiles the players in T cannot predict the partition of $\mathcal{P}_{N \setminus K}$ – in this case, by Expression 2.3, they individually expect the worst. As \mathcal{T} only forms if it is profitable, it will, only, if it is profitable for all $x_i(\tilde{\sigma}, \mathcal{T}, w, (\mathcal{K}, B))$ for all \mathcal{B} . Since $(\tilde{x}, \tilde{\mathcal{P}}) \in C(N, V)$ this is not the case. This, on the other hand implies that the formation of $\tilde{\mathcal{P}}$ is unaffected by possible deviations in this subgame, meeting the first condition of subgame-consistency. \square

Proof of Theorem 1. The proof is by induction. The result holds for trivial, single-player games. Assuming that the result holds for all $k - 1$ player games, the result for k -player games is a corollary of Lemmata 3 & 4. \square

4 Conclusion

Theorem 1 holds for arbitrary games in discrete partition function form, but of course it is most interesting for games where some of the residual cores are empty. When a proposal is made in a game without externalities the invited players do not even (need to) consider the residual game and therefore the emptiness of a residual core is not addressed. Huang and Sjöström (2006) and Kóczy (2009) simply restrict their attention to games where the residual cores are non-empty, in fact the r-core (Huang and Sjöström, 2003) is not even defined for games with empty residual cores. As already pointed out by Kóczy (2007) this is not only an enormous limitation given the number of conditions such games must satisfy (one for each residual game), but the definitions/results do not apply to some games without externalities and so they are not generali-

sations of the well-known results for TU-games. The present paper heals this deficiency.

References

- Asilis, C. M., 1995, A note on the equivalence of time consistency and subgame perfection in stochastic games, *European Economic Review* 39, 245 – 251, special Issue: Symposium of Industrial Organizational and Finance.
- Bloch, F., 1995, Endogenous structures of association in oligopolies, *RAND Journal of Economics* 26, 537–556.
- Bloch, F., 1996, Sequential formation of coalitions in games with externalities and fixed payoff division, *Games and Economic Behavior* 14, 90–123.
- Chander, P. and H. Tulkens, 1995, A core-theoretic solution for the design of cooperative agreements on transfrontier pollution, *International Tax and Public Finance* 2, 279–293.
- Chatterjee, K., B. Dutta, D. Ray, and K. Sengupta, 1993, A noncooperative theory of coalitional bargaining, *Review of Economic Studies* 60, 463–477.
- van Damme, E., 1984, A relation between perfect equilibria in extensive form games and proper equilibria in normal form games, *International Journal of Game Theory* 13, 1–13.
- Eyckmans, J. and H. Tulkens, 2003, Simulating coalitionally stable burden sharing agreements for the climate change problem, *Resource and Energy Economics* 25, 299–324.
- Fershtman, C., 1989, Fixed rules and decision rules : Time consistency and subgame perfection, *Economics Letters* 30, 191 – 194.
- Funaki, Y. and T. Yamato, 1999, The core of an economy with a common pool resource: A partition function form approach, *International Journal of Game Theory* 28, 157–171.
- Huang, C.-Y. and T. Sjöström, 2003, Consistent solutions for cooperative games with externalities, *Games and Economic Behavior* 43, 196–213.

- Huang, C.-Y. and T. Sjöström, 2006, Implementation of the recursive core for partition function form games, *Journal of Mathematical Economics* 42, 771–793.
- Hyndman, K. and D. Ray, 2007, Coalition formation with binding agreements, *Review of Economic Studies* 74, 1125–1147.
- Kóczy, L. Á., 2007, A recursive core for partition function form games, *Theory and Decision* 63, 41–51.
- Kóczy, L. Á., 2009, Sequential coalition formation and the core in the presence of externalities, *Games and Economic Behavior* 66, 559–565.
- Kreps, D. M. and R. Wilson, 1982, Sequential equilibria, *Econometrica* 50, 863–894.
- Kydland, F. E. and E. C. Prescott, 1977, Rules rather than discretion: The inconsistency of optimal plans, *Journal of Political Economy* 85, 473–490.
- Lagunoff, R. D., 1994, A simple noncooperative core story, *Games and Economic Behavior* 7, 54–61.
- Muthoo, A., 1990, Bargaining without commitment, *Games and Economic Behavior* 2, 291–297.
- Muthoo, A., 1995, A bargaining model with players’ perceptions on the retractability of offers, *Theory and Decision* 38, 85–98.
- Osborne, M. and A. Rubinstein, 1990, *Bargaining and Markets* (Academic Press, San Diego).
- Perry, M. and P. J. Reny, 1994, A noncooperative view of coalition formation and the core, *Econometrica* 62, 795–817.
- Ray, D. and R. Vohra, 1997, Equilibrium binding agreements, *Journal of Economic Theory* 73, 30–78.
- Thrall, R. M. and W. F. Lucas, 1963, n -person games in partition function form, *Naval Research Logistics Quarterly* 10, 281–298.

Yi, S.-S., 1996, Endogenous formation of customs unions under imperfect competition: open regionalism is good, *Journal of International Economics* 41, 153–177.